Biometrika (2002), **89**, 4, *pp*. 755–767

Printed in Great Britain

Asymptotic approximations to posterior distributions via conditional moment equations

By JULIE L. YEE

Western Ecological Research Center, U.S. Geological Survey, Sacramento, California 95826, U.S.A.

julie.yee@usgs.gov

WESLEY O. JOHNSON and FRANCISCO J. SAMANIEGO

Department of Statistics, University of California, Davis, California 95616, U.S.A.

johnson@wald.ucdavis.edu samanieg@wald.ucdavis.edu

Summary

We consider asymptotic approximations to joint posterior distributions in situations where the full conditional distributions referred to in Gibbs sampling are asymptotically normal. Our development focuses on problems where data augmentation facilitates simpler calculations, but results hold more generally. Asymptotic mean vectors are obtained as simultaneous solutions to fixed point equations that arise naturally in the development. Asymptotic covariance matrices flow naturally from the work of Arnold & Press (1989) and involve the conditional asymptotic covariance matrices and first derivative matrices for conditional mean functions. When the fixed point equations admit an analytical solution, explicit formulae are subsequently obtained for the covariance structure of the joint limiting distribution, which may shed light on the use of the given statistical model. Two illustrations are given.

Some key words: Bayesian approach; Data augmentation; EM algorithm; Fixed point theorem; Gibbs sampling; Latent data; Screening data.

1. Introduction

The approximation of posterior distributions has a rather long and distinguished history, dating from Laplace (1774). A definitive theory of large-sample approximations to posterior distributions has been primarily built upon the contributions of Le Cam (1953, 1958) and von Mises (1964). Simplifications of the more abstract versions of these theories were published by Walker (1969) and by Chen (1985), while Johnson (1967, 1970) provided higher-order expansions that clarified both the breadth of applicability of the approximation of posteriors by normal distributions and the precise relationship of the error of approximation to the smoothness of the models involved.

This paper is motivated by situations in which the asymptotic behaviour of the posterior distribution is complex and generally unclear. We have in mind problems in which ‘data augmentation’ is possible and where, without such augmentation, the posterior distribution is analytically unmanageable, but where, in the presence of latent data, the posterior distribution can be established as asymptotically normal. As shown by Tanner & Wong

(1987) and Gelfand & Smith (1990), the development of computational techniques involving data augmentation and Gibbs sampling has made it possible to approximate the posterior distributions of these parameters numerically. Our results provide insight into when such posteriors can justifiably be treated as normal.

Let δ be a p-dimensional parameter with probability density function $p(\delta)$, and let Y be observable data generated from the density $p(Y|\delta)$. It may be difficult to handle the posterior density, $p(\delta|Y)$, analytically, while the analysis is greatly simplified by the existence of the q-dimensional vector of latent data, Z, generated from $p(Z|\delta, Y)$. With augmented data (Z, Y), it is often possible to obtain a recognisable posterior density $p(\delta|Z, Y)$; Markov chain Monte Carlo simulation from the full conditional distributions $p(Z|\delta, Y)$ and $p(\delta|Z, Y)$ then gives (δ, Z) pairs that can be used to approximate the joint probability density $p(\delta, Z|Y)$ (Tanner & Wong, 1987; Gelfand & Smith, 1990). Here, we introduce a joint normal approximation for $p(\delta, Z|Y)$ in situations where the full conditional distributions are approximately normal.

We note that the conditions and methods employed for deriving normal approximations are also suitable for approximating joint posterior densities of the form $p(\delta, \phi|Y)$, where the parameter vector is (δ, ϕ), and $p(\delta|\phi, Y)$ and $p(\phi|\delta, Y)$ are approximately normal. The role of ϕ and Z are interchangeable in the development that follows. We restrict our attention exclusively to the latent-data problem.

The approximation to be developed here differs from standard asymptotic approximations in several ways. For example, our asymptotics permit the weight of the prior distribution to grow as the sample size grows, resulting in limiting distributions that may retain some of the effect of one's prior modelling. Also, the asymptotics are pivoted on the solution of certain conditional moment equations.

To contrast our approach with a more standard one, we briefly examine the results obtained by Laird & Louis (1982), a paper that constitutes, to the best of our knowledge, the only published work that treats analytically the normal approximation of posterior distributions in incomplete data problems. Their normal approximation is obtained by matching the mode and the observed Fisher information matrix of the exact and the approximating distributions. They demonstrate the consistency of the approximate posterior and discuss the circumstances under which their approximation provides satisfactory results. We derive an approximating normal distribution under quite different assumptions and by quite different means. In an example in which both methods apply, they are shown to provide very similar results.

Section 2 describes a framework in which approximate normality of $p(\delta, Z|Y)$ is shown to be tenable from asymptotic theory arguments. From this result, an asymptotic normal approximation for $\delta|Y$ is an immediate consequence. Section 3 contains examples and is followed by discussion in § 4.

2. The normal approximation method

2·1. *Definitions, assumptions and preliminary results*

In order to develop our asymptotics, we embed the variables (δ, Y, Z) into a sequence of problems indexed by n; for example we consider (δ_n, Y_n, Z_n), as $n \to \infty$. We thus have sequences of densities $p_n(\delta_n)$, $p_n(Y_n|\delta_n)$, $p_n(\delta_n|Z_n, Y_n)$ and $p_n(Z_n|\delta_n, Y_n)$. For each n, the distributions corresponding to $p_n(\delta_n|Z_n, Y_n)$ and $p_n(Z_n|\delta_n, Y_n)$ are compatible when they define a joint distribution corresponding to $p_n(\delta_n, Z_n|Y_n)$ (Arnold & Press, 1989). The following definitions and assumptions set conditions under which such a sequence of problems will result in a marginal posterior distribution for δ_n that is asymptotically normal.

Since all the analyses that follow are conditioned on the known value of the observed data, Y_n will be treated as a fixed sequence as in Johnson & Gastwirth (1991). We assume that properly normalised data Y_n converge to a limit μ_Y as $n \to \infty$. We define the sequences of conditional mean functions $M_{n,Z}(.)$ and $M_{n,\delta}(.)$, and assume that they converge as $n \to \infty$:

$$M_{n,Z}(s) \equiv E(Z_n | \delta_n = s, Y_n) \to M_Z(s), \quad M_{n,\delta}(t) \equiv E(\delta_n | Z_n = t, Y_n) \to M_\delta(t).$$

For simplicity, we denote the two conditional moments above by $f(s) = M_{n,Z}(s)$ and $g(t) = M_{n,\delta}(t)$ respectively. Our developments rely upon the existence of a point (s^*, t^*) for which $t^* = f(s^*)$ and $s^* = g(t^*)$. The existence and uniqueness of such points are questions that arise in the study of fixed point theorems; see for example Small (1974). We pause briefly to describe the applicable results in this area.

The function f in the preceding paragraph can be pictured as a graph in the (s, t) plane, while the function g can be pictured as a graph in the (t, s) plane. The points of interest are those where these two graphs intersect if drawn together. Equivalently, we are interested in fixed points of the composition mappings, that is values s^* and t^* for which $s^* = g\{f(s^*)\}$ and $t^* = f\{g(t^*)\}$. Existence of such values is guaranteed by the Brouwer Fixed Point Theorem when f and g are continuous functions and the domains of the two compositions are compact, convex sets in a Euclidean space. More general results exist. For example Schauder's Fixed Point Theorem applies to general normed spaces, but the latter are not required by the applications envisaged and will not be discussed here.

The uniqueness of the point (s^*, t^*) above is a delicate matter. The primary applicable result is the Banach Contraction Theorem, which gives a sufficient, but not necessary, condition for uniqueness; in particular it states that any contraction mapping of a complete nonempty metric space into itself has a unique fixed point. In the applications of interest here, uniqueness of the solution (s^*, t^*) is guaranteed if both composition mappings $f \circ g$ and $g \circ f$ are contractions, but these conditions are extremely stringent and would typically be violated. It is nonetheless possible to argue for the uniqueness of the point (s^*, t^*) on other grounds, as will be seen by example in § 2·2. Since the conditions of the Brouwer Theorem hold in the applications we discuss, the existence of solutions (s^*, t^*) is not an issue.

Translating these general remarks into the specific statistical setting introduced above, we assume that a solution exists that satisfies the following system of equations:

$$\mu_{n,Z} \equiv M_{n,Z}(\mu_{n,\delta}), \quad \mu_{n,\delta} \equiv M_{n,\delta}(\mu_{n,Z}). \tag{1}$$

To facilitate asymptotic theory arguments, define the standardisations $\delta_n^* \equiv \sqrt{n}(\delta_n - \mu_{n,\delta})$ and $Z_n^* \equiv \sqrt{n}(Z_n - \mu_{n,Z})$, and note that conditioning upon δ_n or Z_n is equivalent to conditioning upon δ_n^* or Z_n^* respectively.

We require the following additional assumptions.

Assumption 1. The distribution for $(Z_n | \delta_n, Y_n)$ satisfies

$$\sqrt{n}\{Z_n - E(Z_n | \delta_n, Y_n)\} | (\delta_n, Y_n) \to W_1 \sim \mathcal{N}_q(0, \Sigma_Z),$$

in distribution, as $n \to \infty$, where Σ_Z is positive definite.

Assumption 2. The distribution for $(\delta_n | Z_n, Y_n)$ satisfies

$$\sqrt{n}\{\delta_n - E(\delta_n | Z_n, Y_n\} | (Z_n, Y_n) \to W_2 \sim \mathcal{N}_p(0, \Sigma_\delta),$$

where Σ_δ is positive definite.

Assumption 3. The vector functions $M_{n,Z}$ and $M_{n,\delta}$ have finite second derivatives and continuous first partial derivatives,

$$\dot{M}_{n,Z}(s) \equiv \frac{\partial}{\partial \delta'} M_{n,Z}(\delta)\bigg|_{\delta=s}, \quad \dot{M}_{n,\delta}(s) \equiv \frac{\partial}{\partial Z'} M_{n,\delta}(Z)\bigg|_{Z=s},$$

where $\dot{M}_{n,Z}(\delta)$ is a $q \times p$ matrix with (i, j)th element $\partial\{M_{n,Z}(\delta)\}_i / \partial \delta_j$, and $\dot{M}_{n,\delta}(Z)$ is a $p \times q$ matrix with (i, j)th element $\partial\{M_{n,\delta}(Z)\}_i / \partial Z_j$.

Assumption 4. There exists a sequence of solutions $(\mu_{n,\delta}, \mu_{n,Z})$ for the system of equations (1) converging to the limit (μ_δ, μ_Z) so that, in conjunction with the previous assumption, the following limits exist, as $n \to \infty$:

$$\dot{M}_{n,Z}(\mu_{n,\delta}) \to \dot{M}_Z(\mu_\delta) \equiv \dot{M}_Z, \quad \dot{M}_{n,\delta}(\mu_{n,Z}) \to \dot{M}_\delta(\mu_Z) \equiv \dot{M}_\delta.$$

2·2. *Results*

The left-hand side of the expressions in Assumptions 1 and 2 may be expressed in terms of the standardised variables δ_n^* and Z_n^*. Assumption 3 allows us to use a Taylor expansion of the function $M_{n,Z}(\delta_n)$ about $\delta_n = \mu_{n,\delta}$ and of the function $M_{n,\delta}(Z_n)$ about $Z_n = \mu_{n,Z}$. By conditioning on a sequence of (δ_n^*, Y_n) values that converges to an arbitrary value (δ^*, μ_Y) and applying Taylor's approximation, we obtain

$$\begin{aligned}
\sqrt{n}\{Z_n - E(Z_n | \delta_n, Y_n)\} &= \sqrt{n}\{Z_n - M_{n,Z}(\delta_n)\} \\
&= \sqrt{n}\{Z_n - M_{n,Z}(\mu_{n,\delta})\} - \sqrt{n}\{M_{n,Z}(\delta_n) - M_{n,Z}(\mu_{n,\delta})\} \\
&= Z_n^* - \sqrt{n}\{\dot{M}_{n,Z}(\mu_{n,\delta})(\delta_n - \mu_{n,\delta}) + o(\|\delta_n - \mu_{n,\delta}\|)\} \\
&= Z_n^* - \{\dot{M}_{n,Z}(\mu_{n,\delta})\delta_n^* + o(\|\delta_n^*\|)\} \\
&= Z_n^* - \dot{M}_Z \delta_n^* + o(1),
\end{aligned}$$

where the last equality follows from Assumption 4. Thus, conditional on a sequence $\{\delta_n^*, Y_n\}$ that converges to a realisation (δ^*, μ_Y), then, by Assumption 1,

$$(Z_n^* - \dot{M}_Z \delta_n^* | \delta_n^*, Y_n) \to W_1 \sim \mathcal{N}_q(0, \Sigma_Z),$$

in distribution. By a parallel argument, conditional on a sequence (Z_n^*, Y_n) that converges to (Z^*, μ_Y),

$$(\delta_n^* - \dot{M}_\delta Z_n^* | Z_n^*, Y_n) \to W_2 \sim \mathcal{N}_p(0, \Sigma_\delta),$$

in distribution. Thus,

$$(Z^* | \delta^*, Y = \mu_Y) \sim \mathcal{N}(\dot{M}_Z \delta^*, \Sigma_Z), \quad (\delta^* | Z^*, Y = \mu_Y) \sim \mathcal{N}(\dot{M}_\delta Z^*, \Sigma_\delta). \tag{2}$$

It is shown in Appendix 1 that the conditions

$$\Sigma_\delta^{-1} \dot{M}_\delta = \dot{M}_Z' \Sigma_Z^{-1}, \tag{3a}$$

$$(I - \dot{M}_\delta \dot{M}_Z)^{-1} \Sigma_\delta > 0 \quad \text{or} \quad (I - \dot{M}_Z \dot{M}_\delta)^{-1} \Sigma_Z > 0, \tag{3b}$$

where >0 indicates that the corresponding matrix is positive definite, are equivalent to condition (ii) of Theorem 4.1 in Arnold & Press (1989). Since condition (i) of their theorem holds automatically for this problem, the compatibility of the two limiting conditional distributions in (2) is guaranteed. Thus, there exists a joint distribution with these conditionals. The condition in (3a) is necessary for the limiting covariance matrix to be symmetric and will be termed the symmetry condition. We thus obtain the limiting joint normal distribution $(\delta^*, Z^* | Y = \mu_Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, where $\mu' = (0, 0)$ and

$$\Sigma = \begin{pmatrix} (I - \dot{M}_\delta \dot{M}_Z)^{-1} \Sigma_\delta & (I - \dot{M}_\delta \dot{M}_Z)^{-1} \dot{M}_\delta \Sigma_Z \\ (I - \dot{M}_Z \dot{M}_\delta)^{-1} \dot{M}_Z \Sigma_\delta & (I - \dot{M}_Z \dot{M}_\delta)^{-1} \Sigma_Z \end{pmatrix}, \tag{4}$$

which is also derived in Appendix 1. For sufficiently large values of n, we can approximate the distribution for $\delta_n | Y_n$ by $\mathcal{N}\{\mu_{n,\delta}, (I - \dot{M}_\delta \dot{M}_Z)^{-1} \Sigma_\delta / n\}$.

When the solutions $\mu_{n,\delta}$ and $\mu_{n,Z}$ to the system of equations (1) are analytically tractable, the normal approximation for the distribution of $(\delta_n, Z_n | Y_n)$ is in closed form and is thus free from iterative routines. In § 3·1, we present a genetic linkage example as one application for which a normal approximation to the posterior distribution can be derived with only a simple hand calculator. A second illustration involving screening test data, presented in § 3·2, requires equations (1) to be solved numerically using a recursive algorithm similar to the EM algorithm. The method converges very quickly for this problem.

Remark 1. A referee has asked whether or not the compatibility conditions (3) need to be checked since, for each n, we have a proper joint distribution that must necessarily lead to compatible conditionals. The symmetry condition part of compatibility for fixed n requires that the ratio of the conditional densities must factorise as the product of a term involving only δ and a term involving only Z, and it seems reasonable that this might hold in the limit. However, we have been unable to establish this result analytically. For any given problem, it is straightforward to determine that the conditions (3) hold numerically. In our experience, we have always been able to verify the symmetry condition analytically.

Remark 2. While in our presentation we have developed conditional mean equations, we could just as well have focused on conditional mode equations. There will be many situations where there is no analytical form for the conditional means, and consequently the conditional mode is a natural alternative. The corresponding conditional mode equations could then be solved iteratively by a general numerical algorithm. In these instances one could obtain the conditional covariance matrices as the inverses of minus the second derivative matrices for the log conditionals. These matrices would then be evaluated at the solution to the fixed-point mode equations. We believe that it is feasible to develop general-purpose software for implementing this modification of what has been presented here and we are currently exploring this possibility. In Appendix 2, we present a numerical algorithm for obtaining the normal approximation in our second illustration below.

3. Illustrations

3·1. *Genetic linkage example*

We first revisit the genetic linkage problem, which was first treated by Rao (1973, p. 368) and also discussed rather extensively in the data augmentation literature (Dempster et al., 1977; Laird & Louis, 1982; Tanner & Wong, 1987). The problem is identifiable and the estimation of δ using our method can be compared with a method based on obtaining the posterior mode with the expectation-maximisation algorithm in conjunction with the large-sample variance approximation of Laird & Louis (1982), and with a standard Gibbs sampling approach. Our technique can be performed analytically with the aid of only a simple calculator. In both this and the following example, all subscripts n have been suppressed.

The observed data take the form of a multinomial vector,

$$Y \equiv (Y_1, Y_2, Y_3, Y_4)' \sim \mathrm{Mu}\left\{n, \left(\frac{1}{2} + \frac{\delta}{4}, \frac{1-\delta}{4}, \frac{1-\delta}{4}, \frac{\delta}{4}\right)\right\}.$$

The corresponding likelihood in conjunction with a $\mathrm{Be}(a, b)$ prior for δ does not result in a recognisable posterior density for δ, while if the data

$$(Y_1 - Z, Z, Y_2, Y_3, Y_4)' \sim \mathrm{Mu}\left\{n, \left(\frac{1}{2}, \frac{\delta}{4}, \frac{1-\delta}{4}, \frac{1-\delta}{4}, \frac{\delta}{4}\right)\right\}$$

were observed, we would obtain the posterior $(\delta \mid Z, Y) \sim \mathrm{Be}(Z + Y_4 + a, Y_2 + Y_3 + b)$. The latent variable Z, conditional on δ and Y, has distribution $(Z \mid \delta, Y) \sim \mathrm{Bi}\{Y_1, \delta/(\delta+2)\}$, and is standardised as $Z^* = \sqrt{n}(Z/n - \mu_Z)$. From (1), we obtain the finite-n equations

$$\begin{aligned} \mu_\delta = M_\delta(\mu_Z) = E(\delta \mid Z/n = \mu_Z, Y) &= \frac{n\mu_Z + Y_4 + a}{n\mu_Z + Y_2 + Y_3 + Y_4 + a + b}, \\ \mu_Z = M_Z(\mu_\delta) = E(Z \mid \delta = \mu_\delta, Y) &= \frac{Y_1}{n} \frac{\mu_\delta}{\mu_\delta + 2}. \end{aligned} \tag{5}$$

The existence and uniqueness of solutions to the equations (5) merit some commentary. Since the arguments μ_δ and μ_Z are constrained to lie in the closed unit interval $[0, 1]$, and since the functions involved are both rational functions of their arguments and continuous in $[0, 1]$, solutions of these equations are guaranteed to exist in the unit square. Since the two rational functions involved are monotone in $[0, 1]$, there can be but one such solution. Solving for this reveals that there are actually two points in the plane that solve these equations simultaneously, one of them lying outside the unit square and thus being an extraneous solution.

We assume that $a/n \to a^* \geqslant 0$ and $b/n \to b^* \geqslant 0$. As a consequence, the information in the prior need not disappear as the sample size increases. If we assume that $Y/n \to \mu_Y$, as $n \to \infty$, the limits in Assumption 4 exist. The limiting solution (μ_Z, μ_δ) can be obtained analytically and inserted into the derivative functions defined in Assumption 3 to obtain $\dot{M}_Z$ and $\dot{M}_\delta$. The asymptotic conditional variances evaluated at the solutions to the limiting mean equations, Σ_δ and Σ_Z, are

$$\Sigma_\delta = \frac{\mu_\delta(1-\mu_\delta)}{\mu_Z + \mu_{Y_2} + \mu_{Y_3} + \mu_{Y_4} + a^* + b^*}, \quad \Sigma_Z = \frac{\mu_{Y_1}\mu_\delta}{\mu_\delta + 2}\left(1 - \frac{\mu_\delta}{\mu_\delta + 2}\right) = \frac{2\mu_{Y_1}\mu_\delta}{(\mu_\delta + 2)^2}.$$

The symmetry condition in (3a) is verified analytically by showing that $\dot{M}_\delta/\Sigma_\delta = \dot{M}_Z/\Sigma_Z = 1/\mu_\delta$. Analytical results for positive definiteness can also be established with some effect. However they do not appear to add sufficient insight relative to their complexity, so we simply note the positive definiteness for the given data and prior at hand.

Tanner & Wong (1987) presented data with values $n = 197$, $Y_2 = 18$, $Y_3 = 20$ and $Y_4 = 34$. With a uniform prior on δ the solution to (1) is $(\mu_Z, \mu_\delta) = (0{\cdot}1509, 0{\cdot}6240)$. Also, substituting Y/n for μ_y, we obtain $\Sigma_\delta = 0{\cdot}4456$, $\Sigma_Z = 0{\cdot}1150$, $\dot{M}_Z = 0{\cdot}1843$ and $\dot{M}_\delta = 0{\cdot}7141$.

Finally, our normal approximation for the distribution of $\delta|Y$ has mean 0·6240 and standard deviation 0·0511. This is consistent with the EM analysis of Laird & Louis (1982), which produced a normal approximation with mean 0·6268 and standard error 0·0515; Gibbs sampling produced a posterior mean of 0·63 and a standard deviation of 0·05.

3·2. *A screening test problem*

Medical diagnostic tests are performed to determine whether or not a subject has a particular disease or condition. The performance of medical screening tests for a disease is measured by calculating the probabilities of correct diagnoses. The sensitivity of a test is the probability of correctly diagnosing the disease, while the specificity is the probability of correctly indicating that it is not present. These probabilities can be easily estimated by administering the test to patients whose actual disease status is known. The following illustrates a two-test scenario for which there is no gold standard or perfect test available for determining disease status with certainty.

Let the prevalence, or proportion of the disease in the population, be π. Denote the sensitivities of the tests by η_1 and η_2 and the specificities by θ_1 and θ_2, and assume that test outcomes on each unit tested are independent, conditional on disease status, as in Hui & Walter (1980), Joseph et al. (1995) and Johnson et al. (2001). Both tests are conducted on the same set of subjects and their results can be summarised, as in Table 1, as $Y \equiv \{Y_{ij}\}$. Note that $+$ indicates diagnosis of the disease and, for example, the value Y_{12} represents the number of subjects who were diagnosed by the first test as having the disease and by the second test as not having it. The total number of subjects tested is $n = Y_{11} + Y_{12} + Y_{21} + Y_{22}$, and

$$\{Y_{ij}\} \sim \mathrm{Mu}(n, \{p_{ij}\}),$$

with $p_{11} = \pi\eta_1\eta_2 + (1-\pi)(1-\theta_1)(1-\theta_2)$, and so on. There are five parameters and the data are only three-dimensional so an informative prior is required (Neath & Samaniego, 1997).

Table 1. *Two-test data.*

	Test 2	
Test 1	$+$	$-$
$+$	Y_{11}	Y_{12}
$-$	Y_{21}	Y_{22}

Complete and ideal data would consist of the observed table $\{Y_{ij}\}$ and the corresponding table that gives counts for individuals who actually have the disease, $Z \equiv \{Z_{ij}\}$, say. The likelihood based on the augumented data is

$$\pi^{Z_{\cdot\cdot}}(1-\pi)^{n-Z_{\cdot\cdot}}\eta_1^{Z_{1\cdot}}(1-\eta_1)^{Z_{2\cdot}}\eta_2^{Z_{\cdot 1}}(1-\eta_2)^{Z_{\cdot 2}}\theta_1^{Y_{2\cdot}-Z_{2\cdot}}(1-\theta_1)^{Y_{1\cdot}-Z_{1\cdot}}\theta_2^{Y_{\cdot 2}-Z_{\cdot 2}}(1-\theta_2)^{Y_{\cdot 1}-Z_{\cdot 1}},$$

where the dot notation indicates summation over the corresponding index.

Uncertainty about the five parameters, $\delta \equiv (\eta_1, \eta_2, \theta_1, \theta_2, \pi)$, is modelled with independent Beta distributions (Joseph et al., 1995) with respective hyperparameters $(a_{1\eta}, b_{1\eta})$, $(a_{2\eta}, b_{2\eta})$, $(a_{1\theta}, b_{1\theta})$, $(a_{2\theta}, b_{2\theta})$ and (a_π, b_π). Full conditional distributions are independent and Beta; for example, $(\eta_1|Z, Y) \sim \mathrm{Be}(a_{1\eta} + Z_{1.}, b_{1\eta} + Z_{2.})$. The distributions for the elements of Z conditional on (δ, Y) are also independent and binomial; for example, $Z_{11}|\delta, Y \sim \mathrm{Bi}(Y_{11}, \eta_1\eta_2\pi/p_{11})$. Full details of the Gibbs sampler and of the implementation of our approximation are given in Appendix 2.

The components of Y will be large if n is large and if the p_{ij}'s are all bounded away from zero. The full conditional distributions for components of Z are thus approximately normal provided that the corresponding conditional probabilities are also bounded away from zero. If the prevalence, π, were very low or high, our assumptions would not apply. The five independent Beta conditionals will be approximately normal provided their parameters are 'large', which is guaranteed if the hyperparameters of the prior are 'large'. In our experience, a Beta is well approximated by a normal if the interval defined by the mean plus or minus three standard deviations is contained in the unit interval.

We reconsider the Strongyloides infection data discussed in Joseph et al. (1995) and we use the priors elicited by them. The data are $Y_{11}=38$, $Y_{12}=87$, $Y_{21}=2$ and $Y_{22}=35$, and the hyperparameters for the prior are $(a_{1\eta}, b_{1\eta})=(21{\cdot}96, 5{\cdot}49)$, $(a_{2\eta}, b_{2\eta})=(4{\cdot}44, 13{\cdot}31)$, $(a_{1\theta}, b_{1\theta})=(4{\cdot}1, 1{\cdot}76)$, $(a_{2\theta}, b_{2\theta})=(71{\cdot}25, 3{\cdot}75)$ and $(a_\pi, b_\pi)=(1, 1)$. With the help of SAS software, the limiting means, gradients of the means and covariances for δ and Z were obtained numerically, and the compatibility conditions were checked. Details of the complete algorithm are given in Appendix 2. The symmetry condition was checked analytically but details are not given here.

The normal approximation is compared with corresponding Markov chain Monte Carlo results, which were obtained by using WinBUGS version 1.3 software. Table 2 shows that the means and standard deviations obtained from the normal approximation agree well with the Monte Carlo estimates.

Table 2. *Posterior means and standard deviations* (*in parentheses*) *for the Strongyloides infection prevalence. The first two methods use the data and prior information discussed in Joseph et al.* (1995), *while the last two methods use* 10 *times the data and prior hyperparameters*

Method	π	η_1	η_2	θ_1	θ_2
Normal $\times 1$	0·7595	0·8914	0·3074	0·6814	0·9581
	(0·0952)	(0·0394)	(0·0480)	(0·1843)	(0·0210)
Monte Carlo $\times 1$	0·7589	0·8837	0·3103	0·6875	0·9570
	(0·1021)	(0·0419)	(0·0525)	(0·1612)	(0·0214)
Monte Carlo error	$[5{\cdot}4\times10^{-4}]$	$[1{\cdot}5\times10^{-4}]$	$[2{\cdot}3\times10^{-4}]$	$[7{\cdot}4\times10^{-4}]$	$[6{\cdot}0\times10^{-5}]$
Normal $\times 10$	0·7595	0·8914	0·3074	0·6814	0·9581
	(0·0301)	(0·0125)	(0·0152)	(0·0583)	(0·0066)
Monte Carlo $\times 10$	0·7584	0·8909	0·3079	0·6815	0·9580
	(0·0305)	(0·0125)	(0·0154)	(0·0573)	(0·0066)
Monte Carlo error	$[1{\cdot}6\times10^{-4}]$	$[3{\cdot}9\times10^{-5}]$	$[6{\cdot}1\times10^{-5}]$	$[2{\cdot}9\times10^{-4}]$	$[1{\cdot}6\times10^{-5}]$

It is clear, however, that the normal approximation will not be very good for θ_1 in particular since the posterior mean plus two standard deviations is larger than one. We plotted each of the approximations along with the Monte Carlo density approximations obtained by Gibbs sampling and found the fit for η_2 to be very good, the fit for η_1 to be good, the fit for π to be adequate, and the fits for θ_1 and θ_2 to be inadequate because of left skewness. The combined amount of information in the data and the priors is therefore not sufficient for the normal approximation to be adequate overall. However, we also considered the same data and prior information, only multiplied by ten. All plots of exact versus normal approximation were practically identical in this instance. Table 2 also gives means and standard deviations for these data and it is seen that all posteriors are now

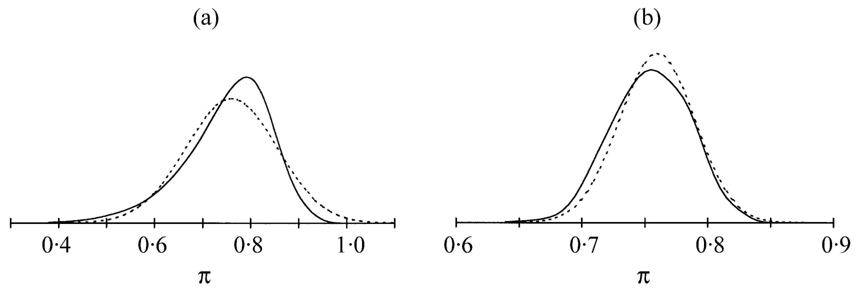

Fig. 1. Monte Carlo (solid) and normal approximation (dashed) estimates for posterior densities of Strongyloides prevalence, π, using (a) 1 times and (b) 10 times the data and prior hyperparameters from Joseph et al. (1995).

well concentrated within the interval (0, 1). Figure 1 gives posterior density estimates for the prevalence, π, with the actual data and prior, and the version with ten times the weight.

Remark 3. A referee was concerned about the adequacy of the normal approximation for this problem when only the sample size was allowed to grow, holding the weight on the prior fixed. To address this issue, we considered a number of situations using versions of the screening data where the sample size was allowed to grow relative to the prior. With the larger sample sizes and fixed prior, the joint posterior becomes increasingly concentrated on a three-dimensional subspace of the five-dimensional parameter space, and this evidently makes Monte Carlo sampling difficult. We generally found that, provided the weight on the prior was not too small relative to that for the data, and that the weight on the prior and the data were sufficiently large, our asymptotics worked well. For example, taking 10 000 times the data, and 1000 or 10 000 times the prior results in easy Gibbs sampling and an excellent normal approximation. With 100 times the prior, Gibbs sampling becomes difficult but, with some effort, we believed the Gibbs sampler had converged and that the normal approximation was very good. Taking a lower weight on the prior resulted in a Gibbs sampler than was extremely difficult and we were not able to verify that the asymptotics were correct. It may be useful to explore the possibility of an alternative parameterisation as in Gelfand & Sahu (1999) for both achieving convergence of the Gibbs sampler and obtaining a good normal approximation. Moreover, Gustafson et al. (2001) devised a special purpose Markov chain Monte Carlo algorithm for a related problem that evidently worked well despite the nonidentifiability. We note that, if a second sample of data were available from a second population with distinct prevalence as in Johnson et al. (2001), then the corresponding statistical model would be identifiable and our asymptotics would apply without regard to the weight on the prior.

4. Discussion

Like the Gibbs sampler, our normal approximation method can be extended to situations involving more than two conditional distributions. One may wonder about the utility of a new method for obtaining asymptotic posteriors at a time when straightforward Markov chain Monte Carlo methods can be used to give results that do not depend on large-sample theory. In the screening example, with an admittedly nonidentifiable model, we found situations where the sample size was large and where obtaining convergence of the Gibbs sampler was difficult, but where our asymptotic methods worked very well. Moreover, the unpublished 2001 University of California, Davis Ph.D. dissertation of

Chun-Lung Su develops these methods further, obtaining asymptotic posterior results for the linear mixed model. Explicit covariance matrices are obtained that shed light on the models and parameterisations considered.

Standard Bayesian asymptotics postulate that the prior distribution is fixed, and that, when supported on the entire parameter space, its influence becomes negligible as the sample size n increases. The present work allows one to derive large-sample approximations in cases in which the prior information deserves substantial weight.

On the negative side, it can be expected that posterior approximations will be poor whenever the posterior mean of any component is less than two standard deviations from a boundary of the parameter space. It is encouraging to see that the normal approximation suggested here, especially with regard to posterior moments, can perform reasonably well for a problem with modest sample size, as demonstrated in the screening example.

Acknowledgement

We thank Chun-Lung Su for his helpful comments and for verifying analytically the symmetry condition for the screening example. We also thank Alan Edelsen for helpful discussions regarding fixed point theory, and two referees for insightful commentary that led to an improved manuscript.

Appendix 1

Discussion of compatibility and its role in the derivation of a joint normal distribution from compatible normal conditionals

Arnold & Press (1989) pointed out that, for compatible conditional distributions, there exists a joint distribution with the corresponding conditionals. They present, in their Theorem 4.1, the equivalence of compatibility and two particular conditions. Their first condition, (i), is automatically satisfied for our problem. The second condition, (ii), involves two parts where the first part, which we have termed the 'symmetry' condition, is assessed as follows. For two conditional densities that are derived from the same joint density, $p_{Y|X}(y|x)$ and $p_{X|Y}(x|y)$, say, we must have $p_{Y|X}(y|x)/p_{X|Y}(x|y) = p_Y(y)/p_X(x)$.

Thus for any given conditional densities to be compatible, the ratio of the conditionals must factorise as the product of separate functions of x and y. The second part of their condition (ii) simply asserts that one of these functions in the factorisation must be integrable.

For normal conditional distributions, the symmetry condition exists if and only if

$$(y-\mu_{Y|X})'\Sigma_{Y|X}^{-1}(y-\mu_{Y|X}) - (x-\mu_{X|Y})'\Sigma_{X|Y}^{-1}(x-\mu_{X|Y}) \tag{A1}$$

is free of xy terms. In the special case where the conditional means are linear combinations of the conditioned variables and the variances are constant, that is $\mu_{X|Y=y} = Ay + E$, $\mu_{Y|X=x} = Bx + F$, $\Sigma_{X|Y=y} = C$, $\Sigma_{Y|X=x} = D$, the xy term in (6) is

$$-2(y'\Sigma_{Y|X}^{-1}Bx - x'\Sigma_{X|Y}^{-1}Ay) = -2y'(\Sigma_{Y|X}^{-1}B - A'\Sigma_{X|Y}^{-1})x = -2y'(D^{-1}B - A'C^{-1})x,$$

which is identically 0 if and only if the symmetry condition, $D^{-1}B = A'C^{-1}$, holds. Note that this can be written as $B'D^{-1} = C^{-1}A$.

Next, note that

$$p_{X|Y}(x|y)/p_{Y|X}(y|x) \propto \exp\{-0{\cdot}5(x-\mu_X)'(C^{-1} - B'D^{-1}B)(x-\mu_X)\},$$

which is integrable if and only if $C^{-1} - B'D^{-1}B > 0$. If the conditionals are compatible, this expression must be proportional to $p_X(x)$. Thus, condition (ii) of Arnold & Press (1989) is satisfied if and only if the symmetry condition holds and if $C^{-1} - B'D^{-1}B = C^{-1}(I - AB) > 0$, which holds if and only if $(I-AB)^{-1}C>0$. We have thus established that (3) is equivalent to compatibility for our limiting conditional normal distributions.

We now derive the joint normal distribution under these two assumptions. When the compatible conditional distributions $X|Y \sim \mathcal{N}(AY+E, C)$, $Y|X \sim \mathcal{N}(BX+F, D)$ determine a joint normal distribution of the form

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right),$$

then

$$A=\Sigma_{XY}\Sigma_{YY}^{-1}, \quad C=\Sigma_{XX}-\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}, \quad E=\mu_X-\Sigma_{XY}\Sigma_{YY}^{-1}\mu_Y,$$

$$B=\Sigma_{YX}\Sigma_{XX}^{-1}, \quad D=\Sigma_{YY}-\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}, \quad F=\mu_Y-\Sigma_{YX}\Sigma_{XX}^{-1}\mu_X.$$

If we solve for μ_X, μ_Y, Σ_{XX}, $\Sigma_{XY}=\Sigma'_{YX}$ and Σ_{YY} in terms of A, B, C, D, E and F, it can be verified that

$$\begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} = \begin{pmatrix} (I-AB)^{-1}C & (I-AB)^{-1}AD \\ (I-BA)^{-1}BC & (I-BA)^{-1}D \end{pmatrix},$$

$$\mu_X=(I-\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1})^{-1}(E+\Sigma_{XY}\Sigma_{YY}^{-1}F), \quad \mu_Y=F+\Sigma_{YX}\Sigma_{XX}^{-1}\mu_X.$$

Appendix 2

Details for the screening example

Let a_δ and b_δ denote the vectors of hyperparameters for the priors and assume that $a_\delta/n \to a_\delta^*$ and $b_\delta/n \to b_\delta^*$ as $n\to\infty$, where all limiting components are positive. The full conditional distributions for $(Z_{ij}|Y, \delta)$ are independent and binomial, namely $Z_{ij}|Y, \delta \sim \text{Bi}(Y_{ij}, p_{ij})$, where

$$\{p_{ij}\} \equiv \{p_{ij}(\delta)\} = \begin{pmatrix} p_{11}(\delta) & p_{12}(\delta) \\ p_{21}(\delta) & p_{22}(\delta) \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\eta_1\eta_2\pi}{\eta_1\eta_2\pi+(1-\theta_1)(1-\theta_2)(1-\pi)} & \dfrac{\eta_1(1-\eta_2)\pi}{(1-\theta_1)\theta_2(1-\pi)+\eta_1(1-\eta_2)\pi} \\ \dfrac{(1-\eta_1)\eta_2\pi}{(1-\eta_1)\eta_2\pi+\theta_1(1-\theta_2)(1-\pi)} & \dfrac{(1-\eta_1)(1-\eta_2)\pi}{(1-\eta_1)(1-\eta_2)\pi+\theta_1\theta_2(1-\pi)} \end{pmatrix}.$$

The full conditionals for $(\delta_i|Y, Z)$ are independent and Beta, namely

$$(\pi|Z, Y) \sim \text{Be}(a_\pi+Z_{..}, b_\pi+Y_{..}-Z_{..}), \quad (\eta_1|Z, Y) \sim \text{Be}(a_{1\eta}+Z_{1.}, b_{1\eta}+Z_{2.}),$$

$$(\eta_2|Z, Y) \sim \text{Be}(a_{2\eta}+Z_{.1}, b_{2\eta}+Z_{.2}), \quad (\theta_1|Z, Y) \sim \text{Be}(a_{1\theta}+Y_{2.}-Z_{2.}, b_{1\theta}+Y_{1.}-Z_{1.}),$$

$$(\theta_2|Z, Y) \sim \text{Be}(a_{2\theta}+Y_{.2}-Z_{.2}, b_{2\theta}+Y_{.1}-Z_{.1}).$$

The limiting fixed point equations are

$$\mu_{Z_{ij}}=\mu_{Y_{ij}}p_{ij}(\mu_\delta), \quad \mu_{1\theta}=\frac{a_{1\theta}^*+\mu_{Y_{2.}}-\mu_{Z_{2.}}}{a_{1\theta}^*+b_{1\theta}^*+\mu_{Y..}-\mu_{Z..}}, \quad \mu_{2\theta}=\frac{a_{2\theta}^*+\mu_{Y.2}-\mu_{Z.2}}{a_{2\theta}^*+b_{2\theta}^*+\mu_{Y..}-\mu_{Z..}},$$

$$\mu_\pi=\frac{a_\pi^*+\mu_{Z..}}{a_\pi^*+b_\pi^*+\mu_{Y..}}, \quad \mu_{1\eta}=\frac{a_{1\eta}^*+\mu_{Z_{1.}}}{a_{1\eta}^*+b_{1\eta}^*+\mu_{Z..}}, \quad \mu_{2\eta}=\frac{a_{2\eta}^*+\mu_{Z.1}}{a_{2\eta}^*+b_{2\eta}^*+\mu_{Z..}};$$

the solution is denoted by (μ_δ, μ_Z). The limiting variances for the normalised conditional distributions are

$$\Sigma_{Z_{ij}} = \frac{\mu_{Z_{ij}}}{\mu_{Y_{ij}}}\left(1 - \frac{\mu_{Z_{ij}}}{\mu_{Y_{ij}}}\right), \quad \Sigma_\pi = \frac{\mu_\pi(1-\mu_\pi)}{a_\pi^* + b_\pi^* + \mu_{Y..}}, \quad \Sigma_{1\eta} = \frac{\mu_{1\eta}(1-\mu_{1\eta})}{a_{1\eta}^* + b_{1\eta}^* + \mu_{Z..}},$$

$$\Sigma_{2\eta} = \frac{\mu_{2\eta}(1-\mu_{2\eta})}{a_{2\eta}^* + b_{2\eta}^* + \mu_{Z..}}, \quad \Sigma_{1\theta} = \frac{\mu_{1\theta}(1-\mu_{1\theta})}{a_{1\theta}^* + b_{1\theta}^* + \mu_{Y..} - \mu_{Z..}}, \quad \Sigma_{2\theta} = \frac{\mu_{2\theta}(1-\mu_{2\theta})}{a_{2\theta}^* + b_{2\theta}^* + \mu_{Y..} - \mu_{Z..}},$$

and $\Sigma_Z =$ block diagonal$\{\Sigma_{Z_{ij}}\}$ and $\Sigma_\delta =$ block diagonal$\{\Sigma_{\delta_i}\}$.

The normal approximation for the posterior distribution can be computed by assuming that a_δ/n, b_δ/n and Y/n have approached their limits, and then by following the following steps.

Step 1. Initialise $i = 0$ and $\hat{\delta}^{(0)} = E(\delta)$.

Step 2. Update estimates: $\hat{Z}^{(i)} = E(Z \mid \delta, Y)|_{\delta = \hat{\delta}^{(i-1)}}$, $\hat{\delta}^{(i)} = E(\delta \mid Z, Y)|_{Z = \hat{Z}^{(i)}}$.

Step 3. Check for convergence. Is the Euclidean distance between the current and previous vectors of iterates less than some tolerance? If not, then increment i and repeat Steps 2–3. If so, then $\mu_Z = \hat{Z}^{(i)}$ and $\mu_\delta = \hat{\delta}^{(i)}$.

Step 4. Calculate numerically the moment derivatives

$$\dot{M}_Z = \frac{\partial}{\partial \delta'} E(Z \mid \delta, Y)|_{\delta = \mu_\delta}, \quad \dot{M}_\delta = \frac{\partial}{\partial Z'} E(\delta \mid Z, Y)|_{Z = \mu_Z}.$$

Step 5. Obtain the asymptotic variance–covariance matrices

$$\Sigma_Z = \lim_{n\to\infty} n \operatorname{cov}(Z \mid \delta, Y)|_{\delta = \mu_\delta}, \quad \Sigma_\delta = \lim_{n\to\infty} n \operatorname{cov}(\delta \mid Z, Y)|_{Z = \mu_Z},$$

and check conditions (3).

Step 6. The distribution of $\delta \mid Y$ is approximately normal with mean μ_δ and covariance matrix $(I - \dot{M}_\delta \dot{M}_Z)^{-1}\Sigma_\delta/n$.

References

Arnold, B. C. & Press, S. J. (1989). Compatible conditional distributions. *J. Am. Statist. Assoc.* **84**, 152–6.

Chen, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. R. Statist. Soc.* B **47**, 540–6.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. R. Statist. Soc.* B **39**, 1–38.

Gelfand, A. E. & Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Am. Statist. Assoc.* **94**, 247–53.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

Gustafson, P., Le, N. D. & Saskin, R. (2001). Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics* **57**, 598–609.

Hui, S. L. & Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–71.

Johnson, R. A. (1967). An asymptotic expansion for posterior distributions. *Ann. Math. Statist.* **38**, 1899–907.

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851–64.

Johnson, W. O. & Gastwirth, J. L. (1991). Bayesian inference for medical screening tests: Approximations useful for the analysis of Acquired Immune Deficiency Syndrome. *J. R. Statist. Soc.* B **53**, 427–39.

Johnson, W. O., Gastwirth, J. L. & Pearson, L. M. (2001). Screening without a gold standard: The Hui-Walter Paradigm revisited. *Am. J. Epidem.* **153**, 921–4.

Joseph, L., Gyorkos, T. W. & Coupal, L. (1995). Bayesian estimation of disease prevalence and parameters for diagnostic tests in the absence of a gold standard. *Am. J. Epidem.* **141**, 263–72.

Laird, N. M. & Louis, T. A. (1982). Approximate posterior distributions for incomplete data problems. *J. R. Statist. Soc.* B **44**, 190–200.

LAPLACE, P. S. (1774). Mémoire sur la probabilité des causes par les évènemens. In *Mémoires de Mathématique et de Physique, presentés à l'Académie royale des sciences, par divers savans & lûs dans ses assemblées*, Tome Sixiéme, 621–56.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. of Calif. Publ. Statist.* **1**, 277–330.

LE CAM, L. (1958). Les propriétés asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris.* **7**, 17–35.

NEATH, A. A. & SAMANIEGO, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *Am. Statistician* **51**, 225–34.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

SMALL, D. R. (1974). *Fixed Point Theorems.* Cambridge: Cambridge University Press.

TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.* **82**, 528–40.

VON MISES, R. (1964). *Mathematical Theory of Probability and Statistics*, Ed. H. Heiringer. New York: Academic Press.

WALKER, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. R. Statist. Soc.* B **31**, 80–8.

[*Received September* 2000. *Revised January* 2002]